Triangle percolation in mean field random graphs — with PDE

BALÁZS RÁTH and BÁLINT TÓTH
Institute of Mathematics
Budapest University of Technology (BME)

March 3, 2008

Abstract

We apply a PDE-based method to deduce the critical time and the size of the giant component of the "triangle percolation" on the Erdős-Rényi random graph process investigated by Palla, Derényi and Vicsek in [4], [9].

1 Introduction

In this Note we investigate triangle percolation in the Erdős-Rényi random graph and other related graphs. The model was defined in [4] and [9] by Derényi, Palla, and Vicsek in order to simulate phase transition of overlapping communities in real networks. They also considered the more general case of k-clique percolation, but we restrict our attention to the $k=3$ case.

The Erdős-Rényi random graph $G(N,p)$, defined in [5], is a random subgraph of K_N, the complete graph on N vertices: The edge set of $G(N,p)$ is chosen at random in the following manner: we declare every edge occupied with probability p, otherwise we call the edge vacant. If N is very large and $p=\frac{t}{N}$, it is well-known that the random graph undergoes phase transition: if we consider the size distribution of the connected components, then a giant component will emerge at $t_c=1$ (this is called the critical time in the $k=2$ case). Because of the mean field property of the graph, it is possible to relate the distribution of the size of the connected component of an arbitrary vertex to the total population of a Galton-Watson branching process with a $\lambda=t$ parameter Poisson offspring distribution (see e.g. [7]).

The branching process (and in the $N\to\infty$ limit, the random graph) becomes supercritical when t, the expected number of first generation offspring (i.e. the number of neighbors of the root) exceeds 1. The limiting component-size distribution of the random graph and the density of the giant component can be determined explicitly with the generating function method.

Vicsek et al. generalize this idea to determine the critical time of a different type of phase transition: if we consider the triangle subgraphs of the Erdős-Rényi random graph and declare two triangles connected if they have a common edge, then a different time-scale is needed to see the emergence of a connected triangulated giant component: the branching process method makes sense in the $N \to \infty$ limit if $p = \frac{t}{\sqrt{N}}$, and $t_c = \frac{1}{\sqrt{2}}$ is the critical time of the $k = 3$ case, see [9].

A different approach to arrive at the same results (in the $k = 2$ case) is to view the evolution of the random graph as a stochastic process: if every edge of the graph turns from empty to occupied state with rate $\frac{1}{N}$, independently from one-another then at time t we will see $G(N, p = 1 - e^{-\frac{t}{N}})$ which is asymptotically the same as $G(N, p = \frac{t}{N})$ as $N \to \infty$. It is easy to relate the evolution of the random graph to the mean field stochastic model of coagulation, the Marcus-Lushnikov process, which converges to the solution of the Smoluchowski coagulation equation (with multiplicative kernel), see (in historical order) [2], [3], [1], [6]. A useful way to handle the Smoluchowski equation is by taking its Laplace-transform (which is essentially the same as using generating functions): the transformed differential equation becomes a well-known PDE, the Burgers equation (see [2], [3], [8]), which can be solved explicitly by using the method of characteristics.

The method of branching processes cannot be applied to a slightly modified random graph process, where the dynamics are the same but the initial graph is an arbitrary graph (and not the empty graph as in the Erdős-Rényi case), but the PDE method generalizes to these models both in the $k = 2$ and the $k = 3$ case. In this Note we derive a (non-linear, non time-homogeneous, first order) PDE which describes the evolution of the $N \to \infty$ limit of the triangulated component-size densities of a general mean-field random graph, and give an explicit solution to that PDE, which enables us to calculate the critical time and the size of the giant triangulated component for an arbitrary initial component size-distribution.

2 Definitions

In the rest of this Note, we will look at the random graph processes on the following time-scale: $G(N, p \approx \frac{t}{\sqrt{N}})$, or more briefly $G(N,t)$: vacant edges turn occupied independently with rate $\frac{1}{\sqrt{N}}$. The edge set of $G(N,t)$ is denoted by $E(N,t)$. In the Erdős-Rényi case, the initial graph has no edges, but it can be itself a random graph in the general case.

In order to describe the triangle-structure of the graph, let us define an auxiliary graph, $\hat{G}(N,t)$ with vertex set $E(N,t)$, and two vertices of $\hat{G}(N,t)$ be connected by an edge if the corresponding $e, f \in E(N,t)$ are edges of the same triangle whose third edge is also in $E(N,t)$. We call the connected components of $\hat{G}$ the triangulated components of G. The "size" or "weight" of a triangulated component is its size in $\hat{G}$.

If $e \in E(N,t)$, denote by $S(e,t)$ the size of the triangulated component of e. Let us denote by $\mathcal{C}_n(N,t)$ the number of triangulated components of weight n.

Summing the total weight of components we get the total number of edges: $\sum_{n\in\mathbb{N}} n\cdot \mathcal{C}_n(N,t) = |E(N,t)|$, and we can define

$$m_1^*(t) := \lim_{N\to\infty} \frac{|E(N,t)|}{\frac{1}{2}N^{\frac{3}{2}}} = m_1^*(0)+t \tag{2.1}$$

by the law of large numbers.

If we define $c_n(N,t) = \frac{\mathcal{C}_n(N,t)}{\frac{1}{2}N^{\frac{3}{2}}}$, then it is natural to expect that $\lim_{N\to\infty} c_n(N,t) = c_n(t)$ exists and is a deterministic nonnegative real number for each n and t. Of course, there is a minimal criterion for this to hold: the sequence of initial random graphs must have the property that the limits $\lim_{N\to\infty} c_n(N,0) = c_n(0)$ exist. We need additional assumptions:

The *asymptotic mean field independence property*, or briefly A.M.F.I.P.: Let us choose a subset $E' \subseteq E(N,t)$ with $|E'| \ll N^{\frac{3}{2}}$ (for example E' can be the set of edges connected to $v_0 \in V(G)$), explore the connected components of the edges of E' in $\hat{G}(N,t)$ and denote the edges contained in the explored triangles by $\overline{E'}$. The A.M.F.I.P. is satisfied if the probability distribution of $S(f,t)$, where $f \notin \overline{E'}$ is asymptotically independent from the distribution of the explored components as $N\to\infty$. Note that this property is weaker then the "branching process" property: there asymptotic independence holds even after a smaller exploration step: if $e\in E'$, then the number of triangles that contain e, but are not contained in E' is asymptotically independent of E'.

We also assume that our sequence of mean field initial graphs has the *asymptotic trivial structure property*, or briefly A.T.S.P.: let us define another auxiliary graph, $\tilde{G}(N,t)$. $V(\tilde{G}(N,t))$ consists of the triangles of $G(N,t)$ and two vertices are connected if the corresponding triangles share an edge. The A.T.S.P. means that asymptotically almost surely the connected component of an arbitrarily chosen vertex of $\tilde{G}(N,t)$ is either the (unique) giant component of $\tilde{G}(N,t)$ or a tree.

An immediate consequence of the A.T.S.P. assumption is that $c_n(t) > 0$ only if n is in $\mathbb{O}$, the set of positive odd numbers: if we remove a vertex from a tree in $\tilde{G}(N,t)$, then we remove exactly two edges of $E(N,t)$. Denote by $\mathbb{O}^\infty = \mathbb{O}\cup\{\infty\}$: in the mean field limit, the size of the triangulated component of an edge e is in $\mathbb{O}^\infty$, $S(e,t)=\infty$ means that e is an edge of the triangulated giant component.

Let us define the Laplace transform (generating function) $C(t,x) = \sum_{n\in\mathbb{O}} c_n(t)\cdot e^{-n\cdot x}$ for $t\geq 0$ and $x>0$. $C(t,x) = \sum_{n\in\mathbb{O}^\infty} c_n(t)\cdot e^{-n\cdot x}$, since $e^{-\infty\cdot x}=0$. Denote the partial derivatives of C with respect to t and x by $\dot{C}$ and C', respectively. It is convenient to define $v_n(t) = n\cdot c_n(t)$ and $V(t,x) = \sum_{n\in\mathbb{O}} v_n(t)\cdot e^{-n\cdot x}$, so that $C'(t,x) = -V(t,x)$ holds. Under the assumption of the mean field properties, the function C will satisfy the following PDE:

$$\dot{C}(t,x) = e^{V(t,x)^2 - m_1^*(t)^2 - x} - 2V(t,x)\cdot m_1^*(t)^2 \tag{2.2}$$

After solving the PDE, we will be able to express $c_n(t)$ as a function of the initial data. Let us emphasize that $m_1(t) := \sum_{n\in\mathbb{O}} v_n(t) = V(t,0_+) \neq m_1^*(t)$ for all

t (although we do assume that $m_1(0) = m_1^*(0)$), because the $N \to \infty$ limit and the $\sum_{n \in \mathbb{O}}$ summation are not interchangeable in the supercritical phase: the breakdown of equality indicates the presence of a giant component, because a positive portion of edges is missing if we sum the weight of small components:

$$m_1(t) + v_\infty(t) = m_1^*(t) = m_1(0) + t \tag{2.3}$$

where $v_\infty(t) \cdot \frac{1}{2} N^{\frac{3}{2}}$ is, up to leading order, the weight of the giant component of $\hat{G}(N,t)$.

3 Derivation of the PDE

In order to derive (2.2), we need some more definitions.

Let us orient the edges of K_N in an arbitrary way, so that we can talk about the "initial" and "final" endpoints of each edge. If e is the new edge that we are about to occupy at time t, with initial and final endpoints u and v, then the "vicinity" of e can be described by a two-variable function $\rho_e^t : \mathbb{O}^\infty \times \mathbb{O}^\infty \to \mathbb{N}$, where $\rho_e^t(i,j)$ is the number of vertices w such that w, u and v form an "(i,j)-type cherry": both $\{u,w\}$ and $\{v,w\}$ are in $E(N,t)$, moreover $S(\{u,w\},t) = i$ and $S(\{v,w\},t) = j$. $i = \infty$ or $j = \infty$ is an admissible choice, because we want to take into account those edges in the vicinity of e that belong to the giant triangulated component. When e becomes occupied, all the components in the vicinity of e merge into one component of size $1 + \sum_{i,j} \rho_e^t(i,j) \cdot (i+j)$, since the merged non-giant components are distinct by the A.T.S.P. The value of $\mathcal{C}_n(t)$ changes by

$$\mathbb{I}[\, n = 1 + \sum_{i,j} \rho_e^t(i,j) \cdot (i+j) \,] - \sum_i \rho_e^t(i,n) - \sum_j \rho_e^t(n,j)$$

We can give the probability distribution of ρ_e^t for an arbitrary e when $N \to \infty$ using the A.M.F.I.P. For each w,

$$\mathbb{P}(S(\{u,w\},t) = i) \approx \frac{i \cdot \mathcal{C}_i(N,t)}{\binom{N}{2}} = \frac{v_i(t)}{\sqrt{N}},$$

up to leading order. Also $\mathbb{P}(S(\{u,w\},t) = \infty) \approx \frac{v_\infty(t)}{\sqrt{N}}$. Using similar estimates and the A.M.F.I.P., the probability that w, u and v form an (i,j)-type cherry is $\frac{v_i(t) \cdot v_j(t)}{N}$, up to leading order. When $N \to \infty$, the number of (i,j)-type cherries in the vicinity of e has Poisson distribution and their joint distribution is the product measure: for any fixed $\mathbf{r} : \mathbb{O}^\infty \times \mathbb{O}^\infty \to \mathbb{N}$, the probability of the event $\{\forall i \forall j \; \rho_e^t(i,j) = \mathbf{r}(i,j)\}$ (or briefly $\{\rho_e^t \equiv \mathbf{r}\}$) is

$$\prod_{i,j \in \mathbb{O}^\infty} e^{-v_i(t) v_j(t)} \frac{(v_i(t) v_j(t))^{\mathbf{r}(i,j)}}{\mathbf{r}(i,j)!} = e^{-m_1^*(t)^2} \prod_{i,j \in \mathbb{O}^\infty} \frac{(v_i(t) v_j(t))^{\mathbf{r}(i,j)}}{\mathbf{r}(i,j)!}$$

We can now start to derive the differential equation (2.2).

Between t and $t+\mathrm{d}t$, approximately $\frac{N^{\frac{3}{2}}}{2}\mathrm{d}t$ edges become occupied, and their contributions to the change of $\mathcal{C}_n(N,t)$ are independent again by the A.M.F.I.P., so we may use the law of large numbers to describe the evolution of the component-size vector:

$$\mathcal{C}_n(N,t+\mathrm{d}t)-\mathcal{C}_n(N,t)\approx\mathbb{E}\left(\mathcal{C}_n(N,t+\mathrm{d}t)-\mathcal{C}_n(N,t)\right)\approx$$
$$\sum_{\mathbf{r}}\left(\mathbb{I}[1+\sum_{i,j}\mathbf{r}(i,j)\cdot(i+j)=n]-\sum_{i}\mathbf{r}(i,n)-\sum_{j}\mathbf{r}(n,j)\right)\cdot\mathbb{P}(\rho_e^t\equiv\mathbf{r})\frac{N^{\frac{3}{2}}}{2}\mathrm{d}t$$

If we divide both sides by $\frac{N^{\frac{3}{2}}}{2}\mathrm{d}t$, let $N\to\infty$ and $\mathrm{d}t\to 0$ and take the Laplace-transform of both sides, then the left-hand side becomes $\dot{C}(t,x)$. Let us calculate the Laplace-transform of the right-hand-side. The first term is

$$\sum_{n\in\mathbb{O}^\infty}e^{-n\cdot x}\sum_{\mathbf{r}}\mathbb{I}[1+\sum_{i,j}\mathbf{r}(i,j)\cdot(i+j)=n]\cdot\mathbb{P}(\rho_e^t\equiv\mathbf{r})=$$
$$\sum_{\mathbf{r}}\mathbb{P}(\rho_e^t\equiv\mathbf{r})e^{-\left(1+\sum_{i,j}\mathbf{r}(i,j)\cdot(i+j)\right)\cdot x}=e^{-m_1^*(t)^2-x}\sum_{\mathbf{r}}\prod_{i,j}\frac{\left((v_i(t)e^{-ix})(v_j(t)e^{-jx})\right)^{\mathbf{r}(i,j)}}{\mathbf{r}(i,j)!}=$$
$$e^{-m_1^*(t)^2-x}\prod_{i,j}\sum_{r=0}^{\infty}\frac{\left((v_i(t)e^{-ix})(v_j(t)e^{-jx})\right)^{r}}{r!}=e^{-m_1^*(t)^2-x}\prod_{i,j}e^{(v_i(t)e^{-ix})(v_j(t)e^{-jx})}=$$
$$e^{V(t,x)^2-m_1^*(t)^2-x}$$

The second term:

$$\sum_{n\in\mathbb{O}^\infty}e^{-n\cdot x}\sum_{\mathbf{r}}\sum_{i\in\mathbb{O}^\infty}\mathbf{r}(i,n)\mathbb{P}(\rho_e^t\equiv\mathbf{r})=\sum_{n\in\mathbb{O}}e^{-n\cdot x}\sum_{i\in\mathbb{O}^\infty}\mathbb{E}(\rho_e^t(i,n))=$$
$$\sum_{n\in\mathbb{O}}e^{-n\cdot x}\sum_{i\in\mathbb{O}^\infty}v_i(t)v_n(t)=m_1^*(t)V(t,x)$$

The third term is handled in the same way.

Putting these equations together we arrive at (2.2). It is convenient to use the shorthand notation $W(t,x):=e^{V(t,x)^2-m_1^*(t)^2-x}$ for the Laplace transform (generating function) of the size of the component we get by occupying a vacant edge at time t. Note that $W(t,0_+)<1$ indicates that this probability distribution is defective in the supercritical case, since $m_1^*(t)-V(t,0_+)=v_\infty(t)>0$.

4 Solution of the PDE

It is possible to give an explicit solution to (2.2) with the method of characteristics. Differentiating the PDE with respect to x and rearranging the equation we get a first order PDE for V:

$$\dot{V}+V'\cdot\left(e^{V^2-m_1^*(t)^2-x}\cdot 2V-2m_1^*(t)\right)=e^{V^2-m_1^*(t)^2-x}\qquad(4.1)$$

Let us consider the following ODE with initial condition $\mathbf{x}(0) = x$:

$$\dot{\mathbf{x}}(t) = e^{V(t,\mathbf{x}(t))^2 - m_1^*(t)^2 - \mathbf{x}(t)} \cdot 2V(t,\mathbf{x}(t)) - 2m_1^*(t) \tag{4.2}$$

If we define $\mathbf{v}(t) = V(t,\mathbf{x}(t))$, then $\mathbf{v}(0) = V(0,x)$. Putting (4.1) and (4.2) together we get a system of differential equations that can be solved without knowing $V(t,x)$ in advance:

$$\begin{cases} \dot{\mathbf{x}}(t) = e^{\mathbf{v}(t)^2 - m_1^*(t)^2 - \mathbf{x}(t)} \cdot 2\mathbf{v}(t) - 2m_1^*(t) \\ \dot{\mathbf{v}}(t) = e^{\mathbf{v}(t)^2 - m_1^*(t)^2 - \mathbf{x}(t)} \end{cases}$$

In order to solve these equations explicitly, define $\mathbf{w}(t) = e^{\mathbf{v}(t)^2 - m_1^*(t)^2 - \mathbf{x}(t)} = W(t,\mathbf{x}(t))$.

$$\dot{\mathbf{w}}(t) = \mathbf{w}(t) \cdot (2\mathbf{v}(t)\dot{\mathbf{v}}(t) - 2m_1^*(t) - \dot{\mathbf{x}}(t)) = 0$$

Thus $\mathbf{w}(t)$ is constant: $W(t,\mathbf{x}(t)) = W(0,x)$, $\mathbf{v}(t)$ is linear: $\mathbf{v}(t) = V(0,x) + t \cdot W(0,x)$, and $\mathbf{x}(t)$ is quadratic:

$$\mathbf{x}(t) = x + (V(0,x) + t \cdot W(0,x))^2 - V(0,x)^2 - (m_1(0) + t)^2 + m_1(0)^2 \tag{4.3}$$

If we start with $\mathbf{x}(0) = 0$, then $\mathbf{x}(t) \equiv 0$ and $\mathbf{v}(t) = m_1(0) + t = m_1^*(t)$, but we know that $V(t,0_+) \neq m_1^*(t)$ if $t > T_g$. This breakdown of analiticity is due to the intersection of characteristics: another characteristic curve $\mathbf{x}(t)$ starting at x intersects the $\mathbf{x}(t) \equiv 0$ curve at the time when (4.3) becomes zero: the intersection time t solves the following equation:

$$t^2 \cdot \left(W(0,x)^2 - 1\right) + t \cdot (2V(0,x)W(0,x) - 2m_1(0)) + x = 0$$

If we let $x \to 0$ in this equation, the solution will converge to $t = T_g$, the first time when another characteristic curve hits 0. We have to divide all the coefficients by x, use $m_1(0) = m_1^*(0) = V(0,0) = \sum_{n \in \mathbb{O}} v_n(0)$, $m_2(0) = -V'(0,0) = \sum_{n \in \mathbb{O}} n \cdot v_n(0)$ and $W(0,0) = 1$ to get -2 times the following equation as $x \to 0$:

$$T_g^2 \cdot (2m_2(0)m_1(0) + 1) + T_g \cdot (m_2(0) + m_1(0) \cdot (2m_2(0)m_1(0) + 1)) - \frac{1}{2} = 0 \tag{4.4}$$

As a special case, if we start from the empty graph, then $m_1(0) = 0$ and $m_2(0) = 0$, thus we get $T_g = \frac{1}{\sqrt{2}}$, which agrees with the critical time obtained in [4] and [9], by using the branching process method. If $G(N,0)$ is uniformly chosen from all graphs with $m_1(0)\frac{1}{2}N^{\frac{3}{2}}$ edges (where $m_1(0) < \frac{1}{\sqrt{2}}$), then this graph is asymptotically the same as the Erdős-Rényi graph at time $t = m_1(0)$, thus $m_2(0) = \frac{m_1(0)}{1 - 2m_1(0)^2}$ and $T_g = \frac{1}{\sqrt{2}} - m_1(0)$. This result can also be obtained by the branching process method.

An example of a sequence of initial random graphs that have the A.M.F.I.P., but do not have the branching process property: let $G(N,0)$ be chosen uniformly from all triangle-free graphs that have $m_1(0)\frac{1}{2}N^{\frac{3}{2}}$ edges. In this case $m_2(0) = m_1(0)$, and

the T_g of this graph is greater than the T_g of the Erdős-Rényi graph with the same $m_1(0)$, but smaller than that of the Erdős-Rényi graph with the same $m_2(0)$. This follows from the fact that the solution of the equation (4.4) decreases if we increase $m_1(0)$ or $m_2(0)$.

In order to express the value of $v_\infty(t)$, let us define $\hat{X}(t,w)$, the inverse function of $W(t,x)$ in the x variable and $\hat{V}(t,w) = V(t,\hat{X}(t,w))$. $\hat{X}(t,w)$ is well-defined and is a decreasing function of w on the interval $(0,W(t,0)]$. Since $w(t)$ remains constant along the characteristics, $\hat{V}(t,w) = \hat{V}(0,w) + tw$ and

$$\hat{X}(t,w) = \hat{X}(0,w) + (\hat{V}(0,w)+tw)^2 - \hat{V}(0,w)^2 - (m_1(0)+t)^2 + m_1(0)^2$$

is expressed explicitly given the initial data. $W(t,0)$ is the smallest w such that $\hat{X}(t,w) = 0$, and $v_\infty(t) = m_1^*(t) - m_1(t) = \hat{V}(0,1) + t - \hat{V}(0,W(t,0)) - tW(t,0)$.

Acknowledgement: The research work of the authors is partially supported by the following OTKA (Hungarian National Research Fund) grants: K 60708 (for B.R. and B.T.), TS 49835 (for B.R.).

References

[1] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.

[2] E. Buffet and J.V. Pulè. On Lushnikov's model of gelation. *Journ. Stat. Phys.*, 58:1041–1058, 1990.

[3] E. Buffet and J.V. Pulè. Polymers and random graphs. *Journ. Stat. Phys.*, 64:87–110, 1991.

[4] I. Derényi, G. Palla, and T. Vicsek. Clique percolation in random networks. *Phys Rev. Lett*, 94:160202, 2005.

[5] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[6] N. Fournier and J-S. Giet. Convergence of the Marcus-Lushnikov process. *Methodol. Comput. Appl. Probab.*, 6(2):219–231, 2004.

[7] S. Janson, T. Łuczak, and A. Rućinski. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[8] G. Menon and R. L. Pego. Approach to self-similarity in Smoluchowski's coagulation equations. *Comm. Pure Appl. Math.*, 57(9):1197–1232, 2004.

[9] G. Palla, I. Derényi, and T. Vicsek. The critical point of k-clique percolation in the Erdös-Rényi graph. *J. Stat. Phys.*, 128:219–227, 2007.

Address of authors:

Institute of Mathematics
Budapest University of Technology
Egry József u. 1
H-1111 Budapest, Hungary

e-mail:
`rathb@math.bme.hu`
`balint@math.bme.hu`